

DRAFT

**HUMAN STUDIES/WORLDWIDE SCIENTIFIC AFFAIRS :
DATA MANAGEMENT AND ANALYSIS UNIT PROPOSAL**

PM3003565362

DRAFT
HUMAN STUDIES/WORLDWIDE SCIENTIFIC AFFAIRS :
DATA MANAGEMENT AND ANALYSIS UNIT PROPOSAL

1) PURPOSE

To develop a data management and analysis unit that will support the mission of Human Studies (HS) to provide understanding on the science of smoking and health, communicate with the scientific and public health community, evaluate proposed product changes, and guide reduced harm product development by developing the capacity to:

- Implement procedures that will guide the receiving, storing and the management of data
- Develop, as necessary, and implement the appropriate methodologies to analyze data
- Analyze and report the results of analysis of data received or generated by Human Studies .
- Present and publish statistical findings of the data

2) RATIONALE

To carry out its mission, Human Studies has developed several research projects aimed at examining cigarette smoke uptake from conventional and electrically heated cigarettes. To conduct these studies HS currently contracts with various Contract Research Organizations (CROs). These organizations will collect data and provide HS with the raw data in a SAS format (without subject identifiers). The CROs will analyze the data specifically to address the objectives as outlined in the study protocols. Further analysis of the data prompted by this initial examination will be done only on a limited basis, if at all, by the CRO with, more than likely, the initiation of additional contracts. These data, however, provide a rich source from which many questions of interest that pertain to the mission of the department can be explored. With the appropriate personnel, hardware and software in place, these additional detailed exploratory analyses can be performed by an in-house data management and analysis unit. Such a unit can:

- Aid in providing the background information needed to guide the development of future studies pertaining to cigarette smoke constituents (i.e., variability estimates for sample size determination)
- Be used to develop methodology to analyze non-linear data such as those from biomarker measures and compound mixtures such as cigarette smoke constituents
- Be used to model relationships of cigarette smoke uptake and explanatory variables of interest such as **dose**,
- Allow for the comparison of data across studies
- Be used as input into the development of future potentially reduced exposure products

PM3003565363

3) SOURCES OF DATA

Potential sources of data include current and future research conducted by Human Studies (HS) as well as studies conducted by external researchers. At present, these sources include the following studies:

- Pilot Study for the Total Exposure Study
- Total Exposure Study
- Electrically Heated Cigarettes
- Japan Oasis Study
- Fading Study
- Cigarette Burn Time

Additional studies that are being carried out by other WSA components maybe included in the database to expand the range of analyses that can be conducted

Data maybe received in a number of formats including existing data files (SAS, Microsoft Excel, etc.) or be entered directly into a database (Microsoft Access, SAS). (see Section ____). All data will be received in a format that will minimize security breaches. Measures that will be taken to enhance security restrictions are detailed in Section ____.

While the databases from these studies contain varying information, certain elements common to each study may provide for comparisons between studies: demographics, number of cigarettes, "dose", biomarker measures, etc.

4) HOUSING AND MAINTENANCE OF DATABASES

The databases will be resident on a server that will be maintained by Information Technology. The server is a _____. Having the data resident on the server lends itself more readily to multi-user applications and will also aid in maintaining the security of the database. Data on the servers are routinely backed up and archived. With the continued growth in the number of studies being conducted by HS, the use of a stand-alone server allows for the anticipated study expansions and volume of data.

5) ACCESSIBILITY

Access to the database will be available only to those persons who are directly involved in the entering or manipulation of data for analysis and presentation. These persons will require approval authority from Bettie L. Nelson, coordinator of Biostatistics/Epidemiology. Each person will review and sign the standards of operation regarding safeguarding the database system (see Section ____) to indicate that they have read it, understand it and will comply to it. Any requests for information from the databases will be made in writing and submitted. The request will be discussed to

6) SOFTWARE

Software currently available in HS that can be used in the management and analysis of data include, the following: SAS System Version 8, StatXact4, LogXact4, STATA6, Microsoft Excel, Comprehensive Meta Analysis, and Microsoft Access. Additional software will be purchased as needed to carry out the appropriate analysis of the data.

7) POTENTIAL USES OF DATABASE

Each study conducted by HS has specific objectives outlined in the protocol and the data analyses are aimed at addressing the specific objectives. However, the completed analysis of the data to address the stated objectives invariably raises other questions that may lead to the consideration of analyses that were not included as part of the study protocol. As the need for these additional supplementary examinations arises these can be addressed through use of the

Additional Personnel (Job Descriptions)

The studies that are being conducted or planned for HS contain a fair number of variables. The Total Exposure Pilot Study contains data from 133 subjects and includes data from the Case Report Form; Enrollment Questionnaire; and Weekly Surveys in addition to . Preliminary plans for the main TES indicate approximately 6,000 subjects maybe enrolled.

To maximize the use of these databases, additional persons will be required to manage and analyze the data for
Data Base Manager

- Senior Research Scientist
- Post-doctoral fellow
- Programmer

Backup and recovery

1. What database product is used (Oracle or Microsoft's SQL)?
2. Who manages the database?
3. Who has access to the database? Safeguards in place?
4. Who makes use of the data from the database and how?
5. Storage capacity? Is this an issue?
6. Back-ups – IT
7. Database Manual of Procedures/Operations?
8. How do you accommodate your end users? (Access, SAS)
9. If SAS, how?
10. How are requests made of the data (form, to whom?)
11. What would you do differently if you had to design a database system today?
12. Do you have persons from different projects who may wish to make use of data from different databases?
13. If so, are the databases set up with similar naming conventions and rules for data entry so that one can relate data from the different files?
14. Would it make sense to have data base from Human Studies formatted in a compatible manner if there was a desire to relate data from the two databases together?
15. What is knowledge set?
16. Do you warehouse data for mining or other purposes?

<http://databases.about.com/compute/databases/gi/dynamic/offsite.htm?site=http://www.pilotsw.com/news/data%5Fwhite.htm>

Why we need a unit

Who is responsible

Who will do what

Potential results

How much will it cost?

Murillo, Joe
ext 2623
24th floor of 120
Specimen given
sensitivity of
data

PM3003565367